DOCUMENT RESUME

ED 354 266 TM 019 551

AUTHOR Herman, Joan
TITLE Fostering Collaboration. Annual Report of Network Activities. Project 1.1.
INSTITUTION National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Nov 92
CONTRACT R117G10027
NOTE 52p.
PUB TYPE Information Analyses (070) -- Reference Materials - Bibliographies (131) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Agency Cooperation; *Educational Assessment; Educational Cooperation; Educational Improvement; Elementary Secondary Education; *Evaluation Methods; *Networks; *Research and Development; Shared Resources and Services; Test Construction
IDENTIFIERS Alternative Assessment; *Center for Research on Eval Standards Stu Test CA; *Performance Based Evaluation

ABSTRACT

This report summarizes a variety of major collaborations in which the Center for Research on Evaluation, Standards, and Student Testing (CRESST) of the University of California at Los Angeles participated in Fiscal Year 1992. Most of the collaborations are part of the CRESST's Program One, "Building the Infrastructure for Improved Assessment." These ventures have included establishing and maintaining several networks for development and improvement of assessment, such as the Science-Math Alternative Assessment Network, and participating in networks established by other organizations to support innovation in assessment. The CRESST has participated in national boards and commissions to promote dialogue in practitioner and research communities, and has collaborated in several state and district assessment programs. Attachments include: (1) a 71-item list of CRESST collaborations with agencies; and (2) the agenda from the 1992 CRESST conference "What Works in Performance Assessment," which contains abstracts of 54 conference presentations. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED354266

National Center for Research on
Evaluation, Standards, and Student Testing

Final Deliverable – November 1992

Project 1.1:
Fostering Collaboration

Annual Report of Network Activities

Project Director:
Joan Herman, CRESST/UCLA

U.S. Department of Education
Office of Educational Research and Improvement
Grant No. R117G10027 CFDA Catalog No. 84.117G

Center for the Study of Evaluation
Graduate School of Education
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

TM019551

BEST COPY AVAILABLE

The work reported herein was supported under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

Table of Contents

Introduction 1

Networks Supporting the Improvement of Assessment Practices 2

Participation in National Commissions and Forums 5

CRESST Annual Conference: "What Works in Performance Assessment" 6

Collaboration With States and Districts 6

Attachments to Project 1.1 Deliverable

1. Listing of CRESST collaborations with agencies
2. Agenda from 1992 CRESST conference "What Works in Performance Assessment"
3. Abstracts of presentations at 1992 CRESST conference "What Works in Performance Assessment"

Project 1.1 FOSTERING COLLABORATION
ANNUAL REPORT OF NETWORK ACTIVITIES

Joan Herman

Introduction

CRESST's Program One, *Building the Infrastructure for Improved Assessment*, seeks to assure that the R&D efforts of CRESST and the many other players currently working on assessment reform will have optimal impact on educational policy and practice. To accomplish this goal, the program works to:

- strengthen the network of those currently engaged in assessment initiatives;
- promote collaboration between CRESST R&D projects and those seeking to make changes in assessment.

In the process, these collaborative ventures assure CRESST's continual interaction with the real worlds of policy and practice, enriching our understanding of the critical R&D issues which are faced in those worlds; the perspective and feedback these interactions provide also help to shape the research agenda at CRESST.

This report summarizes a variety of major collaborations in which CRESST engaged during FY92. These included: establishing and/or maintaining several networks to support the development and improvement of assessment practice; participation in and assistance to a number of networks established by others organizations to support innovation in assessment; participation in national boards and commissions considering assessment issues; the hosting of a major national conference to promote dialogue among practitioner, policymaker and research communities; and collaboration in a number of state and district assessment programs. The latter, including efforts in California, Connecticut, Michigan, Ohio, Vermont, San Diego, Los Angeles, Orange and Ventura Counties, are core parts of projects in Programs Two and Three and thus are only briefly listed below.

Networks Supporting the Improvement of Assessment Practices

In this category are continuing collaboration with the Council of Chief State School Officers (CCSSO) and a new joint venture with the Regional Educational Laboratories to improve assessment in math and science.

The Student Assessment Consortium

Established in 1991 with funding from OERI as a joint project between CRESST and CCSSO, the Student Assessment Consortium continues to support working groups in math, science, art, social studies, workforce readiness, and literacy. CRESST has taken a leadership role in the workforce readiness and social studies groups.

The workforce readiness assessment consortium developed a consensus framework based on definitions of workforce readiness and employability articulated by business and industry expectations; national studies by groups such as SCANS and ASTD; state frameworks (Michigan and New York); school district frameworks (principally Fort Worth, Texas); research centers and regional educational laboratories; and universities. The framework is intended to serve as a generalized definition of workforce readiness which various states and other users can adopt or adapt based on specific local needs; it also serves to identify potential targets for joint development of assessments. The framework identifies six basic categories and illustrative elements for each general category; the six categories include: Personal characteristics and attitudes; interpersonal skills; thinking/problem-solving skills; technology; communication; and systems. The consortium also has agreed on principles to guide assessment development, including assumptions and a general prototype development plan. A meeting in October will target a few areas for development and will consider potential assessment strategies for each. Development will then proceed based on the CRESST expert-novice model.

The social studies consortium, under CRESST's leadership, hosted a meeting of states in the spring to review critical skills in the disciplines—economics, geography, history, civics, etc. Researchers, representatives from learned societies, and states participated. Participating states include:

Alabama, Arizona, California, Connecticut, Georgia, Illinois, Iowa, Massachusetts, Nebraska, New Jersey, North Carolina, North Dakota, Texas, Vermont, Virginia, Washington.

Among next steps identified at that meeting was the development of a consensus map of state and national social studies frameworks. CRESST has conducted a review of available frameworks and will be presenting the draft consensus map of primary student outcomes at the next meeting, scheduled in Washington, DC in December. At that meeting the group also will select a subset of identified outcomes to initiate the development of assessment strategies.

In the social studies area, CRESST also is participating in the design team for CCSSO's Secondary Schools Academic Standards Assessment for U.S. History. This group includes the states of Alabama, California, Florida, Hawaii, Illinois, Kentucky, New Jersey, New York and Utah. The group will jointly develop assessments. A two-part assessment strategy is anticipated, with Part A including open-ended items, on-demand performance items, and interview items, while Part B will include group performance exercises, projects and exhibitions, extended response items, portfolios and other exercise types requiring additional development time.

CCSSO also has recently assumed responsibility for the annual large-scale assessment conference in Boulder, previously under the aegis of the Education Commission of the States (ECS). CRESST has been invited to be a formal collaborator in this conference and to be a principal on the planning team.

The Science-Math Alternative Assessment Network

This network is a collaborative venture among all the regional educational laboratories, with CRESST, to improve math and science assessment across the country; it is part of a larger national initiative to improve math and science education. Building on the CRESST Assessments in Practice Data Base and on its protocol, all of the labs are collecting additional innovative assessments in their regions. The exemplars so collected are to be evaluated and, where appropriate, assembled into a resource bank. A training package also will be developed by the group to assure that all regions of the country have access to both assessment resources and

assessment expertise. During FY92, the design plan for this network was formulated and data collection instruments and procedures were developed; data collection also commenced.

Collaboration in Other Networks

During FY92, CRESST also continued its collaboration with other organizations sponsoring innovative assessment networks: These include:

- The Association for Supervision and Curriculum Development (ASCD), which convenes biannually a group of districts across the country; participating districts include:
  - Deer Valley Unified School District, Phoenix, Arizona
  - Colorado Springs School District
  - Littleton Public Schools, Littleton, Colorado
  - Arlington Heights School District, Illinois
  - Urbandale Community Schools, Illinois
  - Northern Trails Area Education Agency, Clear Lake, Iowa
  - Frederick County Schools, Maryland
  - East Irondequoit Central School District, Rochester, New York
  - Portland Public Schools, Oregon
  - Pittsburgh Public Schools
  - Franklin NW Supervisory Union School District, Swanton, Vermont
  - Bellingham Public Schools, Washington
  - Madison Metropolitan School District, Wisconsin
  - Edmonton Public Schools, Alberta, Canada
  - San Diego City Schools
  - Other representatives from selected state education departments
- The National Education Association which brings together 12 of its state affiliates from those states involved in assessment reform; the participating affiliates are from Arizona, California, Colorado,

Connecticut, Florida, Idaho, Kansas, Kentucky, Michigan, Pennsylvania, Vermont, and Virginia;

- The New Standards Project;
- The Coalition of Essential Schools which is involved in the development of an electronic data base of student exhibitions.

Discussions regarding potential collaboration with all the New American Schools Projects currently are underway. Interactions with the Math Science Education Board and with the National Council of Teachers of Social Studies also are on-going.

Participation in National Commissions and Forums

CRESST partners also participated heavily during FY92 in advisory boards and commissions considering national policy issues in assessment. These included:

- California Assessment Program, Technical Advisory Committee
- California Assessment Program, Policy Advisory Committee
- College Board, Board of Trustees
- Math Science Education Board
- National Academy of Education, Panel for the Evaluation of the NAEP Trial State Assessment, Panel on Testing, Panel on Educational Reform
- National Academy of Sciences, Panel on Confidentiality of Data
- National Adult Literacy Survey, Technical Review Committee
- National Council on Education Standards and Testing
- National Education Goals Panel Resource Committees
- National Science Foundation, Human Resource Development Program
- Third International Mathematics and Science Study, U.S. National Steering Committee
- U.S. Congress, Office of Technology Assessment, Committee on Federal Testing Policy

- U.S. House of Representatives, Committee on Education and Labor, Subcommittee on Elementary, Secondary, and Vocational Education
- U.S. Department of Education, Secretary's Advisory Committee on Chapter 1 Testing
- U.S. Department of Labor, Secretary's Commission on Achieving Necessary Skills, Assessment Resource Group

A listing of agencies with whom CRESST collaborated during the year is appended.

CRESST Annual Conference: "What Works in Performance Assessment"

The annual CRESST conference brought together over 300 individuals to UCLA to share progress in alternative assessment. The conference was co-sponsored by the American Federation of Teachers and the National Education Association and attracted practitioners and policymakers from the national, state, district and school levels, as well as assessment researchers and other R&D providers. While CRESST R&D and CRESST validity criteria were featured, the program also provided a forum for others who have been active in assessment policy, research and practice. Plenary sessions synthesized what we know about the paradigms and expectations underlying new approaches to assessment (including cognitive theory, models of student diversity, policymaker expectations, and teacher needs); the equity issues that need to be addressed; and where we've come in addressing technical issues of validity and implementation issues of collaboration, development and use. Smaller group sessions provided more detailed descriptions of progress being made in developing assessments in specific subject areas—math, science, social studies, literacy, multidisciplinary studies, group process, portfolios. Working discussion groups were interspersed to stimulate active discussion and interaction among all conference participants. The full agenda and abstracts of presentations are provided in the attachments.

Collaboration With States and Districts

The project also was involved in planning with a number of states and districts, collaborating with them on the design and implementation of innovative assessment systems. These included:

- California—advice on all aspects of California Assessment Program
- Connecticut—advice on design for 10th-grade statewide assessment, particularly for piloting of group assessment for literature; provision of content assessment model
- Georgia—agreements to confer on design and analysis of assessment problems of mutual interest
- Hawaii—advice to Commission on Performance Standards
- Massachusetts—provision of content assessment model
- Michigan—design and scoring rubric for high school portfolios (in support of workforce readiness goals)
- New Jersey—planning and training for content assessment/social studies assessment
- New Hampshire—discussion of potential points of collaboration in statewide assessment program.
- Ohio—collaboration in statewide secondary science assessment
- Oregon—planning for piloting of statewide assessment using content assessment model
- Vermont—collaboration in design and implementation of statewide portfolio assessment in writing and mathematics
- Conejo Valley School District—planning and training for alternative mathematics assessment
- Cupertino School District—collaborative planning and implementation of portfolio assessments in writing
- Los Angeles Unified School District—joint planning with United Teachers of Los Angeles on innovative districtwide assessment system; design for work America 2000 project
- San Diego School District—collaborative planning, and training and implementation/adaption of content assessment prototypes
- Santa Monica-Malibu Unified School District—collaborative planning of portfolio assessments with selected teachers
- Ventura Country—training and follow-up with 5 districts interested in stimulating alternative assessment

Attachments to Project 1.1 Deliverable

1. Listing of CRESST Collaborations With Agencies
2. Agenda from 1992 CRESST conference "What Works in Performance Assessment"
3. Abstracts of presentations at 1992 CRESST conference "What Works in Performance Assessment"

CRESST Collaborations With Agencies

CRESST has had consultations with the following agencies and organizations during the past 18 months (at a minimum) and indirectly or directly affected hundreds of others.

Agency for Instructional Television
American Association for the Advancement of Science
American Association of School Administrators
American Educational Research Association
American Federation of Teachers
American Psychological Association
Army Research Institute
Association for Supervision and Curriculum Development
California Assessment Program
California Commission on Teacher Education
California Department of Education
California Educational Research Association
California State Legislature
Coalition of Essential Schools
Colorado Department of Education
Connecticut Assessment Program
Council for Education Development and Research
Council of Chief State School Officers
Education Commission of the States
Educational Testing Service
Far West Regional Laboratory for Educational Research
Harvard Project Zero
Higher Education Research Institute
High Scope Educational Research Foundation
House Committee on Education and Labor, Subcommittee on Elementary, Secondary, and Vocational Education
International Association for the Evaluation of Educational Achievement
IOX Assessment, Inc.
Los Angeles Educational Partnership
Los Angeles Unified School District

Math Science Education Board
National Academy of Education
- Panel for the Evaluation of the NAEP Trial State Assessment
- Panel on Performance-Based Testing

National Academy of Sciences
- Panel on Confidentiality of Data
- Panel on Educational Reform
- Mathematical Sciences Education Board

National Assessment Governing Board (1992 Writing Assessment Panel, Geography Consensus Project)
National Assessment of Educational Progress
National Center for Education Statistics
National Computer Systems
National Conference of State Legislators
National Council of La Raza
National Council on Education Standards and Testing
National Council on Measurement in Education
National Education Association
National Education Goals Panel
National Endowment for the Arts
National R&D Centers
- National Research Center on Cultural Diversity and Second Language Learning
- Center for Research on Effective Schooling for Disadvantaged Students
- National Center for the Study of Writing and Literacy
- Center on Organization and Restructuring of Schools
- National Research Center on Student Learning
- National Center on Adult Literacy

National Science Foundation (Evaluation Advisory Committee for Human Resource Development Division)
New Standards Project
North Central Regional Educational Laboratory
Northwest Regional Educational Laboratory
Office of Educational Research and Improvement
Office of Technology Assessment (Committee on Federal Testing Policy)

Ohio Assessment Program
Organization for Economic Community Development
Policy Forum Congressional Staff (Federal Evaluation Policy)
Research for Better Schools
Second International Mathematics Study (IEA)
Senate Committee on Labor and Human Resources
Social Action Research Center
Social Studies Assessment Consortium
Southeastern Regional Vision for Education (SERVE)
State Collaborative on Assessments and Student Standards
Statewide Policy Committee on California Assessment Program
Summa Associates
Survey Research Center, Institute of Social Research, University of Michigan
The Los Angeles Music Center
The Urban Institute
United States Agency for International Development
United States Department of Education—Chapter 1 Advisory Committee on Testing
United States Department of Energy (Evaluation Policy)
United States Department of Labor—Secretary's Commission on Achieving Necessary Skills
United States Office of the Secretary of Defense, Training and Technology
University of California

Presenters' Abstracts

CRESST Annual Conference

"What Works in Performance Assessment"

September 10-12, 1992
UCLA Sunset Village Conference Center

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)

Presenters' Abstracts

Thursday, September 10, 1992
2:00 p.m. - 3:30 p.m. ***Assessment Challenges: Changing Views of Learning, Instruction and Assessment***

Robert Glaser

CRESST/Learning Research and Development Center, University of Pittsburgh
Learning and Assessment

New methods in performance assessment will need to rely on modern knowledge of thinking, reasoning, and cognition. Aspects of this knowledge that are influencing conceptions of learning are: the nature of memory, the analysis of problem solving, children's early foundations of competence, self-regulatory abilities, and the access to knowledge afforded by cultural experiences. Based on these features, the conditions of effective learning include assessment of the following: (a) the increasing coherence and integration of knowledge, (b) the proceduralization of knowledge, (c) knowing how to use resources, (d) the opportunities and constraints provided by the prior knowledge of learners, (e) individual differences in abilities for self-regulation of learning, (f) the social display and social modeling of competence and achievement, (g) making the thinking of a learner overt, (h) engagement in holistic situations rather than overly segmented tasks, and (i) ability and disposition to acquire knowledge and skill.

Thursday, September 10, 1992
2:00 p.m. - 3:30 p.m. ***Assessment Challenges: Changing Views of Learning, Instruction and Assessment***
Edmund W. Gordon

CRESST/City University of New York
Nature of Diversity

Group and individual differences in the characteristics with which students come to school have long been a source of concern for educators. Focus has been placed first on differences in status: ethnicity, gender, class, and caste. More recently, attention has turned to differences in function: affective response tendency, cognitive response tendency, identity, culture, and motivation.

There is little clear evidence of relationships between such differences and specific educational treatments, yet interest in the potential of the paradigm persists. A reconceptualization of the problem is offered to reflect the role of dissonance reduction in the efforts at achieving complementarity between learner characteristics and learning experiences in the interest of impacting the learner behaviors which ultimately influence learning outcomes. Implications for the manipulation of learning experiences to shape learner behaviors in instructional and assessment situations are addressed in this paper.

Thursday, September 10, 1992
2:00 p.m. - 3:30 p.m. ***Assessment Challenges: Changing Views of Learning, Instruction and Assessment***
Lorraine McDonnell

CRESST/University of California, Santa Barbara
Policymaker Beliefs About Assessment

Recent events have demonstrated once again that policymakers' views about student assessment shape the form that it takes in local communities. In-depth interviews with national and state policymakers over the past year show that many believe new forms of assessment can act as potent levers for educational change. However, their expectations are quite varied, ranging from the use of tests as symbolic "wake-up calls," to vehicles for curricular coherence, to mechanisms for

imposing major consequences on schools and students. The design of new forms of assessment is often a highly politicized activity in the sense that a wide range of interests and goals must be accommodated within the same system. As a result, the technical task of test development is complicated by the need to integrate these competing purposes into a system already characterized by uncertain technology. Furthermore, although policymakers hold high expectations for new assessment alternatives, they have little information about the costs and feasibility of implementing them. Still, policymakers remain optimistic about the potential of alternative assessments, despite the cautions raised by testing experts and educators.

Those developing and implementing new forms of assessment need to consider seriously the beliefs that undergird recent policies: Policymakers' expectations will significantly influence how assessment information is used and the level of political support accorded schools. Nevertheless, many officials are mindful that they are moving into uncharted terrain and that the success of new assessment initiatives will ultimately depend on their willingness to enhance local capacity. Consequently, while policymakers' beliefs are often strongly-held, they can be altered by new information generated from research and practice.

Thursday, September 10, 1992
2:00 p.m. - 3:30 p.m. ***Assessment Challenges: Changing Views of Learning, Instruction and Assessment***
Charlotte Higuchi

CRESST/Farmdale Elementary School, Los Angeles
Teacher Needs

What could we change today that would directly improve the quality of education? Performance-based assessment is one answer. But the new assessments' emphasis on problem-solving tasks, social, and personal skills must be equaled in complexity and sophistication by a challenging instructional program. As performance-based assessments strive to assess the whole child—intellectual, social, and personal growth—the instructional program must teach the whole child, demanding a major change in how and what teachers teach.

Nationwide improvement of assessments, curriculum, and instructional strategies requires no less than a total restructuring of how school district administration and school site management support teachers and student learning. Teachers need training, a rearrangement of traditional school schedules to allow for more planning/scoring time, collaboration with university-based researchers, a dissemination network to access information from other teachers/researchers, class sizes of 20 students who are in an ungraded setting for a minimum of two years, and support from parents. The list of teacher needs adds up to deep systemic change. Give teachers the responsibility for the change, along with the necessary materials and structural support to enable them to do their job.

Thursday, September 10, 1992
2:00 p.m. - 3:30 p.m. ***Assessment Challenges: Changing Views of Learning, Instruction and Assessment***
Maxine Frost

Riverside (CA) Unified School District, Board of Education
A Local Policymaker Response

Educational accountability is a high priority for the conscientious school board member—especially if that member means to seek reelection.

The process of meaningful evaluation of student progress has undergone many shifts and changes in the 25 years of my school board service in Riverside. However, key elements necessary in learning assessment, from the perspective of a local policy worker, remain. Evaluations must be understandable, they must be useful, and they must be effective.

Thursday, September 10, 1992
3:45 p.m. - 5:00 p.m. ***Assessment and the Equity Challenge***
Monty Neill

National Center for Fair and Open Testing (FairTest)

Equity is defined in this presentation as a process of inclusion. It is not simply the willingness to do well on someone else's terms, but rather involves changing the terms themselves, through a process of negotiation among involved parties.

Implications of equity as inclusion include the following concepts: The definition for excellence is subject to change and negotiation; equity must address race, class, gender, language, and national origin; multi-culturalism must be anti-racist; the school experience must be inclusive. Making schooling and assessment inclusive faces complex problems. These involve: (a) education of teachers and administrators; (b) parent understanding of what schools should do; (c) schools not encouraging students' desire to learn because students are not free to build on their identity; and (d) parents and school personnel conforming to existing standards that are not equitable.

In the area of assessment, we need to consider two main foci: assisting and documenting learning of individuals, and accountability and improvement of systems. We need to think of how the two levels each can help attain equity and ensure that the two are mutually supportive. The presentation considers how assessment itself can be held accountable so that it can help attain equity and excellence in education, and proposes strategies for doing so.

Thursday, September 10, 1992
3:45 p.m. - 5:00 p.m. ***Assessment and the Equity Challenge***
Linda Winfield

CRESST/UCLA

This presentation expands upon an earlier published commentary (Winfield & Woodard, *Education Week*, January 29, 1992) to provide a review for addressing equity and diversity issues in performance-based measures included in current national testing proposals. There are a number of important questions that must be addressed by both researchers and practitioners. Specifically, within most schools and classrooms the antecedent instructional conditions will impact student test performance. In this presentation, questions will be explored, such as: How do we define equitable learning opportunities? What is the relationship between instructional validity and performance-based outcomes? What are indicators of equitable "school delivery" standards?

Thursday, September 10, 1992
3:45 p.m. - 5:00 p.m. ***Assessment and the Equity Challenge***
Lily Wong Fillmore

University of California, Berkeley
Equity and Fairness in Assessment: The Case of Asian-Americans

In this presentation, I comment on equity and fairness issues in assessment as they relate to Asian American students. Unlike other language minority students, Asian-Americans as a group generally perform well in standardized academic achievement measures, leading observers to characterize them as a "model minority," a group that does well despite the language and cultural barriers that are said to prevent other groups from making progress in school. I will discuss evidence suggesting that there is more—or perhaps less—here than meets the eye. Their test scores are often not an accurate measure of how well or poorly they are doing in school, and as a consequence, many Asian-Americans do not get the educational services and attention they need in school. I will consider these issues especially as they relate to assessments of language and literacy skills, and I will discuss how assessments in these areas figure in the allocation of instructional services that are provided to students with special needs.

Friday, September 11, 1992
8:30 a.m. - 10:00 a.m. ***The Assessment Challenge: The National Scene***
Michael Feuer

Office of Technology Assessment, U.S. Congress
Standards and Assessment: The Congressional Role

The past two years have witnessed unprecedented Congressional involvement in the public debate over educational standards and testing. Several House and Senate bills have specific standards-setting and assessment provisions, and pending reauthorizations of OERI, Chapter 1, Vocational Education, and the ESEA will keep assessment issues high on the debate agenda. The session will review and interpret the recent Congressional session with respect to the evolution of standards and assessment legislative initiatives. In particular, participants will: highlight the positions of their respective political parties; explain areas of agreement and of continued tension; describe the implications for practitioners and scholars of

alternative legislative initiatives; and preview the kind of issues that are likely to come up in the new Congress.

Friday, September 11, 1992
8:30 a.m. - 10:00 a.m. ***The Assessment Challenge: The National Scene***
Andrew Hartman

Committee on Education and Labor, U.S. House of Representatives
Twists and Turns with Standards and Assessments

The recent history of Congress' role in the debate about voluntary national education standards and assessments will be described. Part of this presentation will present the origins of the debate and the political context including the Administration, Governors, and other political forces. Both the House and Senate have passed legislation dealing with these issues. Of special interest is the House provision on "school delivery standards" which deal with equity of opportunity to learn. The political and educational implications and complications of this provision of the House bill will be a focus of the talk. Finally, the role that the academic community has played in the development of this public policy will be described.

Friday, September 11, 1992
10:15 a.m. - 11:45 a.m. ***What We Know: Small Group Sessions–Math***
Tom Romberg

National Center for Research in Mathematical Sciences Education
Performance Assessment in Mathematics: Issues and Status

My purpose in this presentation will be to discuss nine issues being faced by the mathematical sciences education community as attempts are being made to develop an authentic examination system for school mathematics. The issues are presented and discussed in Romberg and Wilson (in preparation) and include: underlying assumptions about the nature of mathematics; underlying assumptions about the learning of mathematics; the need for new psychometric models; alignment with the reform curriculum; specification of performance standards; developing authentic tasks; measuring status, growth, or a combination; scoring—by whom and in what form; and, making reports of results understandable to the public. As each issue is presented, the status of current work related to that issue will be summarized.

Friday, September 11, 1992
10:15 a.m. - 11:45 a.m. ***What We Know: Small Group Sessions–Math***
Suzanne Lane

Learning Research and Development Center, University of Pittsburgh
Empirical Evidence for the Reliability and Validity of QUASAR's Performance Assessment

With the increasing use of performance assessments that measure long-term educational outcomes and growth, evidence is needed to ensure valid assessments of students' proficiency. This presentation will provide both logical and empirical evidence for the reliability and validity of a performance assessment (QUASAR Cognitive Assessment Instrument; QCAI) designed to measure students' complex thinking and reasoning skills in mathematics. The QCAI is designed to measure long-term educational outcomes and growth, and to evaluate whether the goals for the instructional programs are reached. Cognitive significance (Glaser, 1989), content comprehensiveness (Frederiksen & Collins, 1989), and task and rater generalizability (Linn, Baker, & Dunbar, 1989) are but a few areas that will be addressed in the presentation.

Friday, September 11, 1992
10:15 a.m. - 11:45 a.m. ***What We Know: Small Group Sessions–Multidisciplinary***
Pamela Aschbacher

CRESST/UCLA
Multidisciplinary Assessments

There is growing interest across the nation in interdisciplinary education. But how will we assess student achievement in such programs, and how can such assessment inform instruction? Over the past two years we have been exploring the value of using student portfolios as part of a formative evaluation of an interdisciplinary program in the humanities at the secondary level in Los Angeles (Humanitas). Specifically, we are interested in portfolios as a window on both student achievement and the experienced interdisciplinary curriculum. We are also interested in how portfolio assessment can help teachers improve interdisciplinary instruction. Since assignments help define how students spend their time, what they perceive as valued by the teacher, and what the teacher expects of the

students, assignments are a critically important aspect of any program. Portfolios offer an indirect way of obtaining samples of both teachers' assignments and student work.

Raters in our study successfully scored assignments on several dimensions of task quality aligned with Humanitas Program goals, including the extent to which they require students to use complex reasoning, make interdisciplinary connections, make authentic connections to the world beyond the school, self-evaluate, work in cooperative groups, and express themselves in non-written work. In addition, raters looked at student achievement of key program goals, including personal investment in class work, interdisciplinary connections, complex reasoning, self-evaluation and growth over time. Data from six high school classes or teams also included surveys, interviews, class observations, and school records. The findings suggest that portfolios provide a useful measure of the interdisciplinary curriculum and student work. Our work has also provided guidelines for teachers to create assignments that truly achieve their interdisciplinary instructional goals.

Friday, September 11, 1992
10:15 a.m. - 11:45 a.m. ***What We Know: Small Group Sessions–Multidisciplinary***
Daisy Vickers

North Carolina State Department of Public Instruction
Multidisciplinary Assessments

This year the North Carolina State Department of Public Instruction will implement a new, high stakes, state-developed, end-of-grade test. The format will be census testing at grades three through eight in reading, math, social studies, and science and will include both multiple-choice and open-ended items that require extended student responses. In addition, hands-on science and math items will be included. This presentation will cover a discussion of the development, field-testing, and scoring of the open-ended items. Attention will be focused on aligning the assessment with a state-mandated curriculum and with the integration of the content areas. Results will be presented of a study comparing reader agreement and reader rates of a group of hired scorers and classroom content teachers who read and scored the student responses to the open-ended field tests. Statistics on the reliability of scorers will also be presented. In addition to the end-of-grade open-ended testing, the challenges faced in the last school year when a literature-based

English II essay test was implemented on a statewide basis as an end-of-course test will be discussed along with a new, grade-three foreign language listening assessment and a new high school foreign language proficiency writing and speaking assessment.

Friday, September 11, 1992
10:15 a.m. - 11:45 a.m. ***What We Know: Small Group Sessions–Multidisciplinary***
Charles Jervis

Auburn High and Middle Schools, Montgomery Co., Virginia
Multidisciplinary Assessment: Art in Biology

The presenter will give an overview of his efforts to incorporate multidisciplinary instruction into science classes. Emphasis will be upon art alternative assessment in biology classes. Alternative assessments which are multidisciplinary tap into student strengths and increase opportunities for student success. By demonstrating connections across the disciplines, students get a better picture of a holistic educational process. Although sometimes time-consuming and more demanding on teacher preparation time initially, multidisciplinary assessments give a more complete view of student thought and performance ability. The teacher has found that there is a notable increase in student enthusiasm and participation in alternatives when they are presented and that multidisciplinary assessments with interdepartmental cooperation offer a meaningful and realistic alternative to traditional assessments.

Friday, September 11, 1992
10:15 a.m. - 11:45 a.m. ***What We Know: Small Group Sessions–Group Assessment***
Richard Durán

University of California, Santa Barbara
Qualitative Assessment of Learning Interaction

How do students construct and guide their learning in cooperative learning activities? Qualitative analysis of students' cooperative learning interaction using methods of sociolinguistics and discourse analysis reveals students' use of interactional and learning performance strategies that are not well-articulated by educators nor captured by existing static forms of performance assessment.

Analysis of interaction reveals ways in which students negotiate their understanding of academic task demands and participate in mutual assessments of their learning performance with and without teacher inputs and assessments. This presentation discusses evidence of such findings based on videotaped observations of language minority and non-language minority students participating in a language arts curriculum known as Cooperative Integrated Reading and Composition. Discussion is also given to selective use of qualitative assessments of classroom videotaped interaction as supplements to other forms of assessment. Videos of children's interaction over the course of the school year could be included in a formative assessment portfolio. Trained teachers could present cooperative learning videos to parents and other school staff for the purpose of documenting children's academic development over the school year.

Friday, September 11, 1992
10:15 a.m. - 11:45 a.m. ***What We Know: Small Group Sessions–Group Assessment***
Noreen Webb

CRESST/UCLA
Group Assessment

Large-scale assessment programs are increasingly turning to group assessment in which small groups of students collaborate to solve problems or complete projects. Reasons for using group assessment include the following: (a) to reflect the growing importance being placed on cooperative learning and group problem-solving in classroom instruction, (b) to measure students' collaborative skills and problem-solving processes, (c) to provide complex, realistic problems in authentic assessment that may be less intimidating to students in collaborative settings, and (d) logistical reasons, such as making more efficient use of limited test materials.

Little is known, however, about the validity of data from group assessment, especially the extent to which inferences can be made about individual students' competence from group performance data. The study reported here compared performance in small-group and individual assessment contexts and examined the following questions: (a) How well does performance in group problem-solving contexts represent the skills of individual students? (b) What additional information about students' skills is provided by data on group dynamics and group problem-solving processes? Two seventh-grade general mathematics classes at an urban

middle school worked in heterogeneous small groups to calculate the costs of long distance telephone calls (e.g., "A long distance call to San Francisco costs $0.30 for the first minute plus $0.08 for each additional minute. What is the cost of a 10-minute call?"). Students submitted their papers. All groups were tape recorded for the entire class period. Two weeks later, after a review session, students worked on similar problems individually without collaborating with others. The results showed that performance from a group assessment context in which students are allowed to collaborate in their written work is not a valid indicator of students' competence when tested individually. Performance from the group context (average of 94% correct) greatly overestimated students' competence on the individual test (average of 62% correct). The data on group processes showed why students performed so much better in groups than in individual assessment. When students did not understand how to solve the problem, or made errors, other members of the group simply supplied the correct procedures so that their papers would be correct. These results suggest that group assessment may demonstrate what students can do when collaborating with others, but it is not valid for making inferences about the individual competence of students.

Friday, September 11, 1992
10:15 a.m. - 11:45 a.m. ***What We Know: Small Group Sessions–Group Assessment***
Marvin Awbrey

Fresno Unified School District/California Assessment Program
History-Social Science Group Assessment in California (High School Level)

The presentation will focus on the History-Social Science group performance tasks that have been developed and field tested in California high schools through the California Assessment Program. Each of the tasks includes group and individual student activities which yield both group and individual student results. The committee believes that assessment should model good instructional techniques. What we have learned from Group Performance Assessment Tasks: Small groups should include no more than five or six students to encourage active participation of all students; each packet of documents should be limited to no more than six brief sources (i.e., political cartoon, article, map, speech excerpt); students should not be used as the evaluators for the group performance tasks; three evaluators are needed to assess each group performance task and should include the classroom instructor; teachers in the field test liked the group performance task process and including

both group and individual student assessment activities; "scramble" activities developed to briefly model the process of the larger task have been used with teachers and would be appropriate for student use; group performance tasks need to be designed to involve students of various ability levels by including cartoons, maps, and other graphics; the draft scoring guide for group performance assessment tasks includes four categories using a 6-point scale: Group and Collaborative Learning (20%), Critical Thinking (30%), Communication of Ideas (20%), and Knowledge and Use of History (30%); each new group performance task developed brings new challenges and questions.

The Grade 10 History-Social Science CAP Advisory Committee will continue to develop and field test group performance tasks that are aligned with the Grade 10 course "World History, Culture, and Geography: The Modern World (1789 to Present)" in the *1987 California History-Social Science Framework*.

Friday, September 11, 1992
10:15 a.m. - 11:45 a.m. ***What We Know: Small Group Sessions–NAEP Update***
Roy Truby

National Assessment Governing Board
Update on the National Assessment of Educational Progress

This presentation will provide an overview of the National Assessment of Educational Progress (NAEP) assessment in 1992 and 1994. NAEP has historically been a pioneer in innovative assessments. Since its inception in 1969, NAEP has used direct measures—actual writing samples, matrix-sampling, and student background questions. NAEP has more recently broken new ground by including such performance assessments and creative projects as: a hands-on task in science and mathematics; a literacy task involving newspapers, documents, and reference materials; oral and written responses to literature; dynamics of group decision-making; and, student portfolios. The presentation will highlight the new techniques in the 1992 NAEP assessment in mathematics, reading and writing. The 1994 NAEP assessments include mathematics, science, reading, U.S. history, and geography.

Friday, September 11, 1992
10:15 a.m. - 11:45 a.m. ***What We Know: Small Group Sessions–NAEP Update***
Lee R. Jones

Educational Testing Service
Update on the National Assessment of Educational Progress: Incorporating Hands-on Science Tasks Into a Large-Scale National Assessment

The National Assessment of Educational Progress (NAEP) is the U.S. Department of Education's ongoing assessment of what American fourth-, eighth-, and twelfth-grade students know and can do in basic subject areas. An entirely new NAEP science assessment is currently being developed for national administration in 1994. This innovative assessment includes a performance-based, hands-on component for all students. Throughout 1992, extensive school-based pilot testing of 17 different 20- and 30-minute tasks has been conducted in preparation for national-level field testing in February 1993. This discussion will focus on: (a) the philosophical framework and assessment specifications that guided the development of the new NAEP science assessment; (b) the strategies used to develop hands-on tasks that strive to assess students' scientific investigative skills, practical reasoning abilities, and conceptual understanding of science within a relatively short time period; (c) the general results of the pilot testing and the lessons learned about developing hands-on science tasks; and (d) plans for administering and scoring the tasks in next year's national-level field test.

Friday, September 11, 1992
10:15 a.m. - 11:45 a.m. ***What We Know: Small Group Sessions–Service Delivery***
Leigh Burstein

CRESST/UCLA
Service Delivery Standards

Delivery standards are seen by many as essential for successful educational reform. They are a mechanism for ensuring that new assessment systems and student performance standards will be accompanied by improvements in instruction and schooling. And, they would provide a needed contextual framework for interpreting student achievement results. But delivery standards are less understood and more controversial than content standards (e.g., NCTM) because they raise the specter of

nationally dictated educational practices that are better left to local control. Nevertheless, if it were possible to operationalize and measure standards for the learning opportunities and instructional conditions that schools should provide all children, the information provided by such measures could contribute to more informed discourse about the status and progress of children within the nation's schools.

Developing delivery standards is in many respects akin to developing performance standards. Three relevant questions that will likely undergird such an effort are: (a) How should we go about developing delivery standards? (b) How do we validate the comparability, equity, and credibility of delivery standards? (c) How should information about delivery standards be reported?

To date, we have little experience in developing service delivery standards and need to explore alternative models for their development. Substantive and methodological decisions about delivery standards will likely be guided by four pertinent literatures: (a) modern conceptions of teaching, schooling, and curriculum; (b) recent work in the area of teacher assessment; (c) the Education Indicators literature on monitoring the quality of schooling experiences, learning opportunities, and equity; and (d) modern evaluation literature on monitoring innovations and reforms.

Friday, September 11, 1992
10:15 a.m. - 11:45 a.m. ***What We Know: Small Group Sessions–Conative Assessmen***
Richard Snow

CRESST/Stanford University
Conative Aspects of Assessment

The conative aspect of student performance concerns motivation and volition for learning from instruction as well as for performing in assessment situations. The conative category includes such psychological constructs as achievement motivation, anxiety, effort investment, mindfulness, interest, self-regulation, action control, and related personal learning styles. This presentation discusses the importance of conative differences among students in both instruction and assessment and the difficult problems involved in assessing them. A particular concern is conation in

relation to new forms of assessment. Progress in research on these problems is reviewed. Suggestions for further research and for practice are considered.

Friday, September 11, 1992
10:15 a.m. - 11:45 a.m. ***What We Know: Small Group Sessions–Conative Assessment***
Jacqueline Cheong

Far West Laboratory for Educational Research and Development/University of California, Berkeley
Conative Aspects of Assessment

This small group session will present a brief synthesis of progress in assessing cognition. Particular emphasis will be placed on the influence of technologies such as performance assessments on what is assessed, how assessment is done, and how information gathered as a result of assessment is used. Examples from the California Assessment Program and the California Learning Record will be shared.

Friday, September 11, 1992
1:45 p.m. - 2:30 p.m. ***What Can Alternative Assessment Really Do for Us?***
H.D. Hoover

University of Iowa

In this presentation a number of positive effects of an increased emphasis on performance measures will be noted. These will center around classroom assessment and include: a more balanced view of students' achievements when reported to parents and the public; an increase in the measurement knowledge of teachers and school administrators; a more comprehensive curriculum; and improvements in instructional methods. It will also be argued that increased recognition by professional organizations such as NCTM and NCTE of the important role testing plays in student learning may possibly be the most important outcome of the alternative assessment movement.

In addition, a number of educational problems that performance assessment cannot be expected to solve will be discussed. Foremost among these will be the contention that the effects from the misuse of tests brought about by top-down accountability

programs will not be alleviated by the use of performance measures and may in fact be exacerbated.

Friday, September 11, 1992
2:30 p.m. - 4:00 p.m. ***What We Know: Technical Considerations***
Eva L. Baker

CRESST/UCLA
Design and Scoring Performance Assessments

How can performance assessments be designed? Do we have knowledge or hunches that can help us improve the design of procedures in use for new performance-based assessments? This presentation will discuss a strategy for preparing performance assessments to optimize their utility for instructional improvement and reporting purposes. The approach simply suggests that there are reasonable templates that can be used to guide task development for efficiency and potentially for more instructional effectiveness. A key element of the approach is the identification of scoring rubrics by using expert-novice approaches. The rubric focuses on the cognitive demands of tasks rather than on the details of content. The rubric is public. The rubric is applied to particular topics and subject matters, and specific training is provided for each different topic.

Friday, September 11, 1992
2:30 p.m. - 4:00 p.m. ***What We Know: Technical Considerations***
Robert L. Linn

CRESST/University of Colorado at Boulder
Comparing Performance Assessments

Comparability is a fundamental issue whenever assessments are used for high-stakes purposes such as the certification of individual achievements. Statistical equating and calibration procedures used with current standardized tests may or may not be applicable to some of the comparability problems raised by performance assessments. At best, however, those procedures will deal with only limited aspects of the comparability problem. Five approaches (equating, calibration, statistical moderation, prediction, and social moderation) to linking assessments to achieve different types of comparability will be distinguished. These five approaches will be

related to Haertel's (personal communication) framework for thinking about three types of comparability of performance assessments. These are comparability when (a) the same tasks are used in different locations, (b) alternative tasks are intended to be used interchangeably to assess the same skills in the same content domain, and (c) tasks differ in terms of content domain or required skills but performances are judged to be of comparable worth.

Friday, September 11, 1992
2:30 p.m. - 4:00 p.m. ***What We Know: Technical Considerations***
Desmond Nuttall

University of London
Moderation: Lessons From Performance Assessments in the United Kingdom

The British have never taken a strong line towards test equating, relying rather on the judgment of experts to ensure that standards of assessment are (reasonably) constant over time, over examining bodies and between schools and colleges. Despite research pointing to some imperfections in many of the systems in use (e.g., Schools Council Examinations Bulletin 37), there is considerable public confidence in the moderation of performance assessment and an absence of litigation about examination and test results.

The paper describes some of the methods currently in use, and the training given to moderators. The importance of training across a range of material and contexts is stressed (given the research evidence about interaction between moderators and topics), as is the importance of retraining and keeping assessment and moderation skills up to date. The value of moderation as a device for professional training and updating in subject and teaching skills is also considered.

Friday, September 11, 1992
2:30 p.m. - 4:00 p.m. ***What We Know: Technical Considerations***
Richard Shavelson

CRESST/University of California, Santa Barbara
Designing Validity Studies

To validate means to bring logical and empirical evidence to bear on a proposed interpretation of a measurement. In the current testing reform, to design validation studies presupposes a working definition of "achievement" in a subject matter. This paper points out the diverse and limited conceptions of achievement that predominate. It proposes some dimensions of a working definition of achievement that can be used to raise validity questions. Finally, the paper brings data to bear from a series of studies on performance-based assessment, drawing examples from elementary mathematics and science. The specific questions/studies addressed are: impact of content specifications, sensitivity to domain-specific knowledge, sensitivity to instructional history, and exchangeability of different measurement methods.

Friday, September 11, 1992
4:15 p.m.-5:30 p.m. ***What We Know: Models for Collaborative Assessment Development***
Daniel Resnick

CRESST/Learning Research and Development Center, University of Pittsburgh
The New Standards Project

The New Standards Project is a major research and development enterprise with state, district, and emerging university/private sector partners. It is a process- and product-oriented partnership collaboration that has learned any number of lessons about enhancing collaboration. Among them are lessons related to motivation, the need for large and small scale collaboration; the power of prior shared knowledge and experience; the interplay between careful planning and future flexibility; and needs for integration with state and district goals and for multiple and inter-related targets of collaboration. Who needs to be involved? Collaboration intended to change current practice requires a "buy-in" at many levels: teachers, principals, parents, students, subject area specialists and supervisors, assessment heads, chief

state school officers and their staffs, school boards and superintendents, state legislatures, governors.

Friday, September 11, 1992
4:15 p.m. - 5:30 p.m. ***What We Know: Models for Collaborative Assessment Development***
Lorrie Shepard

CRESST/University of Colorado at Boulder

The Colorado project designed to help third-grade teachers develop classroom performance assessments in reading and mathematics will be described. Research questions address the effects of performance assessments on teacher thinking, on instructional practices, and on student learning, and the credibility of these assessments for school board members and parents. A commitment was made to support teachers in developing their own assessments in hopes that this would ensure teacher ownership and better prospects for integrating assessment and instruction. Teachers have expressed a need to start small, for example, tackling new forms of assessment and scoring criteria for only one goal in reading and one in mathematics for the first quarter. The level of technical and instructional support required does not bode well for extending the project to many classrooms. However, in the second year we will see whether the examples and expertise created in the district can extend the use of performance assessments with only workshop support from the university. If nothing else, our collaboration is making teachers more articulate about what knowledge and support they need to implement performance assessments.

Friday, September 11, 1992
4:15 p.m. - 5:30 p.m. ***What We Know: Models for Collaborative Assessment Development***
Daniel Koretz

CRESST/RAND Corporation

RAND has been collaborating with the state of Vermont in the development and evaluation of the state's portfolio assessment program for more than four years. In the development stage, RAND staff were among a sizable number of individuals

with diverse expertise called upon from time to time to provide ideas and to critique the emerging system. In contrast, the evaluation of the program—focusing on implementation and impact, as well as measurement quality—has been primarily RAND's responsibility from the outset.

A number of positive and negative aspects of the Vermont model of collaboration are becoming apparent. The collaborative model of development permitted a strong feeling of ownership by educators and probably contributed to its staff development potential. On the negative side, however, it has permitted decisions that were undesirable in terms of measurement quality. The Vermont model of collaboration in evaluation appears to be providing useful formative information, has constrained the use of the assessment data when it was of insufficient quality, and appears to have increased support for the program in some local districts. The success of this model, however, may depend on the expectations of the Vermonters and their unusual public commitment to evaluation of their program.

Friday, September 11, 1992
4:15 p.m. - 5:30 p.m. ***What We Know: Models for Collaborative Assessment Development***
Thomas Payzant

San Diego City Schools

The district is involved in a number of collaborative models in the development of performance assessment. Dr. Payzant will discuss the guiding principles behind assessment development in the district and then highlight a few of those collaborative models. He will conclude with a summary of next steps in collaboration.

Guiding Principles for Assessment Development: Assessment is considered an integral part of teaching and learning, not an external, isolated activity; assessment development is tied to the district's restructuring efforts in the area of teaching and learning; assessment development centers on the primacy of teachers in the development processes and places increased value on teacher judgment; assessment develop promotes equity for all students to be able to demonstrate what they know and can do; assessment development is based on one high, consistent standard for all students.

Examples of Collaborative Models: National–National Alliance, New Standards Project, Harvard PACE, Urban Math Collaborative, ongoing collaboration with Pittsburgh, Rochester (NY), and White Plains, Pacesetter, NBPTS; State–Collaboration with CAP, AB 40 (California Assessment Collaborative), California Learning Record; Foundations and Universities–Stuart Foundation and CRESST; Local–Individual school and teacher efforts, magnet programs, proposed new assessment system.

Next Steps in Collaboration: The district is working towards the use of performance assessment across schools, projects, and subject areas. We are looking for resources in a number of areas to help us accomplish these next steps.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Social Studies***
David Niemi

CRESST/UCLA
Performance Assessment in Social Studies

What are the implications of performance assessment models for instruction? How can performance assessment inform instructional design and teaching practice? Are performance assessments sensitive to instruction? To investigate ways that assessment research might support instructional improvement, and to test the sensitivity of performance assessment tasks to instruction, CRESST researchers have introduced high school history and geography teachers to an empirically-validated performance assessment model. Our teacher-researcher consortium has developed and tested instructional activities and materials based on the assessment model. This presentation will focus on steps in the development and refinement of the performance assessment model and scoring rubric, and on results from our studies of the impact of the assessment model on instruction.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Social Studies***
Judith Torney-Purta

University of Maryland

Performance Assessment Using Cognitive Mapping and Problem-Solving Tasks in Social Studies

The use of performance assessment in social studies has been less extensive than in some other subject areas. The projects to be described have used cognitive mapping and problem solving in hypothetical situations to measure students' knowledge structures in three topic areas in world history and global studies: Apartheid, the global environment, and the Holocaust. Two programs for 12- to 16-year-olds have been the sites of the research, one a summer program on international studies and the other, three classrooms for gifted and talented students participating in a museum education project. Two models from the literature on cognition have formed the basis for these measures of individuals' conceptual structure and change. The first model is the work of Voss on problem-solving in the social sciences; the method here, however, used graphic schema maps representing the content of oral responses in think-aloud problem solving in contrast to Voss' analysis of type of reasoning (and was used for think-aloud problem solving concerning Apartheid). The second model is the work of Novak on concept maps, augmented here by a new scoring system (and was used for concept maps of the Holocaust and of the global environment drawn over a period of 2 to 3 weeks). The presentation raises issues concerning the usefulness for performance assessment of measures of students' cognitive structure such as concept maps and hypothetical problem solving. Concrete examples will be given. Planning is underway to adapt software developed at Bank Street and at Kings College (London) to make it possible to collect data from students at computer terminals.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Social Studies***
Cris Gutierrez

Jefferson High School, Los Angeles

How would you defend the Salem witches in an appeal? Metaphorically speaking, how would you assess progress in 20th century America: as an electric chair? a seesaw? What would you say in a Public Service Announcement on AIDS? In this session we'll explore and discuss some examples of students' alternative assessments from Thomas Jefferson High School Humanitas' American Studies course. Let's begin to see how these alternative assessments challenge students' problem-solving and critical thought, as well as address their multiple intelligences, while encouraging their deep understanding of content areas. What are the strengths and strains of such alternative assessments? How can these and others be initiated and refined? What are some appropriate scaffolds—for both students and teachers? Can these alternative assessments really work with inner-city students? (Thomas Jefferson is a large, year-round, inner-city school.)

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Science***
Gail Baxter

University of Maryland
Performance Assessment in Science: Prior Research and Current Directions

Efforts to develop alternative measures of student achievement in science focus on creating assessments that display how students use their knowledge to reason and solve problems in contextually relevant situations. Prior research examined the reliability, validity and exchangeability of these assessments. Results suggest: (a) Raters can be trained to reliably score performance but large numbers of tasks are needed to get a dependable measure of student performance. (b) These alternative assessments measure different aspects of science achievement than do traditional aptitude or achievement measures, and they distinguish students with varying instructional histories. And, (c) measurement methods are not exchangeable. Different methods measure different aspects of achievement. Given some preliminary evidence of the technical quality of these assessments, the

question is: Do these alternative assessments tap higher-order thinking skills (as the developers intended) or are they primarily rote recall of science facts and procedures packaged in a new way? Current research is addressing this question. Efforts focus on describing how performance actually elicited from students in varying assessment situations differs among students at various levels of achievement. Documentation of the link between performance scores and cognitive demands of alternative assessments will provide important information to guide future test development.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Science***
R. Darrell Bock

CRESST/National Opinion Research Center, University of Chicago
Performance Assessment in Science

Extended response exercises such as essay questions and laboratory performance tasks are now considered an important part of secondary school science assessment. In large-scale assessments, the problem of scoring the extended responses is formidable, not only because of the cost of reading the papers, but also because of typically poor inter-rater agreement. In a recent CRESST/NORC field trial of a prototype 12th-grade science assessment, papers from open-ended exercises were scored using as readers Illinois science teachers. Following detailed instructions, the readers scored each exercise by the so-called "graded mark-point method" as well as a global, more impressionistic rating of the student's performance. The graded mark-point method identified the essential issues, concepts, or results that the student is expected to cover in his or her response.

For the mark-point ratings, the agreement of readers was reasonably good, but less so for the global rating. The main difficulty is that, in spite of the well-defined categories, some readers do not use any of the extreme categories while others do. We now believe that it may be necessary to make adjustments for these response sets when computing respondents' IRT scales.

We are not yet satisfied that ratings are sufficiently sensitive to students' thinking processes, as opposed to content knowledge. We are presently investigating the

possibility of refining the global rating to assess dimensions of conceptual understanding and effectiveness of the approach to the problem set by the exercise.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Science***
Kathleen B. Comfort

California Assessment Program, California Department of Education
Performance Assessment in Science

In the past, large-scale assessment in most states has taken the form of standardized, multiple-choice tests that ask students to choose the correct response from among a set of alternatives These tests have many negative characteristics and consequences: They tend to narrow the curriculum, encourage the teaching of disconnected, low level facts, frustrate teachers and students, and undermine school improvement efforts.

Current research suggests that changing the way we measure student achievement may improve the quality of science instruction. In California, the desire for measurable results has focused attention on the need for improvements in how we measure student learning, what we measure, and how that information is used to restructure educational programs (Carlson, 1992) The California Assessment Program (CAP) has begun introducing performance-based assessments that will indicate more directly what students actually know, how well they think, and what they can do. In science, a variety of innovative performance measures are currently under either development, field testing, or implementation. They include performance tasks, open-ended or free response questions, modules, and portfolios. This presentation will focus on the rationale for including performance measures in large-scale science assessment, CAP's experience in the development and field testing of these new measures, and the preliminary results and findings. Samples of performance tasks, open-ended questions, and portfolio guidelines will be available for review and discussion.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Science***
Melody Ulen

Littleton High School, Colorado
Graduation Assessment in Science—Littleton High School

Students graduating from Littleton High School, beginning with the class of 1995, will be required to demonstrate competency in 19 different graduation requirements. The following graduation requirement addresses science: The LHS graduate applies skills and scientific concepts to explain his/her world, find solutions for its problems, and suggest improvements in the quality of life. In order to demonstrate this competency, the student must achieve at least a Proficient rating in each of two tasks. The tasks are, in brief: (a) design, perform, interpret, and discuss a scientific experiment; and (b) read an article involving an environmental problem, research the validity of the problem, and propose solutions, either to the actual problem or to the issue of reporting an invalid problem.

Each of these tasks was field tested at LHS during the 1991-1992 school year. The scoring rubric for each task was tried and problems were identified. Testing conditions for each task were modified as necessary. There are three levels possible on the final score for each task: Excellent, Proficient, and Unacceptable. Students must earn at least a Proficient rating to graduate. Students practice each of these assessments in "Integrated Science" classes which they take to prepare for the final graduation demonstration.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Literacy***
Elfrieda Hiebert

University of Colorado at Boulder
Performance Assessment in Literacy

Each subject area poses unique issues in the development and implementation of performance assessment at the instructional and policy tiers of assessment. Critical dimensions of performance assessments in literacy will be the focus of this presentation. For example, the material that students read needs to be considered

in designing and evaluating performance assessments. Texts differ on critical elements like genre, length, difficulty level, and source (i.e., self-selected, teacher-selected). Examples from state, district, and classroom performance assessments will be given to illustrate the range of decisions for a particular element.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Literacy***
Beth Rogers

Charlottesville City Schools, Virginia
Performance Assessment in Literacy

The assessment of literacy development is the act of learning about the evolving communicative competence of children within the school community. The ways in which teachers learn about students depend upon their knowledge of literacy acquisition and development, their beliefs about teaching and learning, and their values about schooling in general. Teachers who view language acquisition as both a cognitive and social process, who recognize that the communicative arts are interrelated and that literacy events must be meaning-centered and purposeful, establish classrooms in which assessment and instruction mirror one another. In these classrooms assessment means studying children, learning as much as possible about each child. The teacher as the "anthropologist" (Cambourne & Turnbill, 1990) becomes an active member in the community of learners, both transcribing the learning as well as participating in and orchestrating the literacy events within the classroom. In this setting, teachers not only document each child's individual development, they also study the social dimension of learning. This type of assessment is based on the concept of "zone of proximal development," which enables teachers to look carefully at what students can do with assistance and how it affects their future learning (Ellwein, in press; Vygotsky, 1978). Finally, since language learning is context dependent, astute teachers chronicle the effect of task and material on each child's ability to demonstrate competence.

This type of assessment is great in contrast to the traditional role of the teacher, which has been to teach, test, grade, and report with no examination of whether or not these practices are consistent with what is known about child growth and development. As teachers acquire knowledge about literacy acquisition and examine

their practices, beliefs and values, their perspectives on assessment and instruction change.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Literacy***
Barbara Weiss

California Assessment Program, California Department of Education
Performance Assessment in Literacy

Recently 170 California educators came together in Sacramento to score California's new English-language arts assessment—an assessment that invites students to construct their own meaning by bringing personal experiences into transactions with a text, to discuss the text with other students, and to write an essay that may be directly linked or thematically linked to the reading. The response of the classroom teachers to this new assessment and the scoring process was overwhelmingly enthusiastic.

Included in the presentation will be an overview of the holistic scoring guide used to score the reading assessment, examples of student work, and a video that shows the link between classroom practices and the assessment. A brief discussion of the components of the new assessment system for California will also be covered.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Workforce***
Harold F. O'Neil, Jr.

CRESST/University of Southern California
Workforce Readiness Assessment Strategies

This session will provide an overview of the various definitions of workplace readiness, the type of competencies that have been identified in this area, as well as the variety of approaches to assessment that have been and are being developed. The emphasis in this session will be on the outcomes that are general and interdisciplinary in nature, and on assessment strategies that use performance or other non-traditional measures. Work to be developed in the future will also be discussed.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Workforce***
Edward Roeber

Council of Chief State School Officers
Workforce Readiness Assessment Strategies

This session will provide an overview of the various definitions of workplace readiness, the type of competencies that have been identified in this area, as well as the variety of approaches to assessment that have been and are being developed. The emphasis in this session will be on the outcomes that are general and interdisciplinary in nature, and on assessment strategies that use performance or other non-traditional measures. Work to be developed in the future will also be discussed.

Saturday, September 12, 1992
8:30 a.m. – 10:15 a.m. ***What We Know: Small Group Sessions–Portfolios***
From CRESST/UCLA: Maryl Gearhart, Joan L. Herman, Shelby A. Wolf, Eva L. Baker
From UCLA: Geoffrey B. Saxe, Deborah J. Stipek, Susie W. Hakansson

Portfolios as Assessment: The Design of Portfolios in Language Arts and Mathematics

This session will provide brief introductions to two R&D efforts to collaborate with teachers in the design and uses of portfolios for classroom assessment at the elementary level. One project is focused on students' writing, the other on mathematics. In common across the projects are the goals: (a) to ground the design of assessments in developmental research on students' understandings and competencies; (b) to enhance teachers' subject matter and developmental knowledge so that teachers can make informed judgements of students' complex work (and help students learn to do the same); and (c) to design classroom assessment practices that can be realistically and effectively integrated with instruction.

We are focusing on portfolios because they have considerable potential as tools for educational reform. Effective uses of portfolios for assessment, however, require teachers' solid understanding of reform curricula. Teachers need to understand, for

example, a great deal about writing and writing development in order to offer helpful feedback to students as they develop and revise their compositions. The presentation will review staff development activities that are targeted to (a) enhancing teachers' knowledge of these domains and patterns of students' development within these domains; and (b) providing practice with specific assessment techniques. Progress on scale development also will be discussed.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Portfolios***
Paul LeMahieu

Pittsburgh Public Schools
Portfolios as Assessment

Since 1987, the Pittsburgh Public Schools have been engaged in the development of a portfolio process as a way to integrate assessment into the learning process. The resulting portfolios include early drafts, works in process, revisions, reflections, and self-evaluations related to successful and unsuccessful pieces, as well as learning logs and journals. They are designed to give a complete "portrait" of the student as a learner and have evolved into a very powerful tool for organizing and integrating curriculum, instruction, and assessment.

The district has developed an evaluative framework for judging the performance of students, resulting in a complex assessment scheme organized around three major dimensions: Accomplishment as a Writer; Use of Process and Resources; and Engagement, Growth, and Developing as a Writer.

The training of raters and systematic scoring of over 2000 portfolios gave researchers the opportunity to analyze the three broad areas of inquiry: the nature of standards and consistency of judgments, the exploration of consequential and related validities, and the effects of cultural similarities and differences between students and raters. The presentation summarizes relevant analyses, results and conclusions in these areas.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Portfolios***
Willa Spicer

South Brunswick (NJ) Board of Education
Portfolios as Assessment

This discussion will focus on two types of elementary school portfolios: (a) a kindergarten through grade 2 portfolio designed to enable teachers and other school people to assess the progress of children as they move through these grade levels, and (b) a grades 3-12 "Best Works" portfolio designed to help students become better at assessing and improving their own written work. These examples will be used to highlight several important issues in portfolio development: (a) The purposes of data collection determine what should be collected and when it should be collected. (b) The process of developing a portfolio is complex and takes many years. (It may never end.) (c) There are methods of dealing with reliability and accountability which include teachers and enhance the utility and quality of portfolio assessments.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Portfolios***
Mary Barr

University of California, San Diego
Portfolios Plus: The California Learning Record

After a brief overview of the California Learning Record Project, a videotaped illustration of the CLR in use in a K-1 "sheltered English" classroom in San Diego will be shown. Participants will then examine a completed record for one of the children shown in the video. The assessment includes multiple viewpoints about the child's learning and is especially sensitive to bilingual learners.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Reform/Innovation***
Ruth Mitchell

Council for Basic Education and Pelavin Associates
A National Study of Performance Assessment as Education Reform

Assessment of Student Performance is one of 12 studies commissioned by the OERI under the general heading "Evaluating Education Reform." The project has just completed its first year and is intended to take three years. Pelavin Associates, in partnership with the Educational Testing Service and the Council of Chief State School Officers, is looking at performance assessment from the perspective of organizational change. Results from our first year indicate that practitioners have three main concerns about performance assessment: (a) They need information about and experience with these assessments. They are especially interested in seeing videos of assessment in action. (b) They are worried about the technical aspects of performance assessment: standards, reliability and validity, relevance to the curriculum, equity, uses and purposes. (c) They realize that the switch to performance assessment demands resources now in short supply.

During a national conference, it became obvious that the core issue in performance assessment is the quality of teaching. In discussion after discussion, the focus shifted from the assessment itself to the need for changes in teaching revealed by the assessment results. Our organizational change focus will enable us to research the impact of performance assessment not only on teachers but also on the other components of the educational system.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Reform/Innovation***
Bruce McIntosh

Edmonton Public Schools, Alberta, Canada
School Reform and Innovative Assessment

Edmonton Public Schools monitors the performance of its schools and the district relative to results in the following areas: student performance, student attitudes, staff performance, staff attitudes, parent attitudes, community attitudes, physical

assets, and finances. The assessment of student achievement and performance is the focus of developmental activities aimed at enhancing current practices. The district has determined that these practices will include annual measures for every student based on the total curriculum. Measures will utilize teacher evaluations of student performance as the basis for the information being reported.

Assessment strategies being developed utilize student performances of assigned tasks that are aligned with provincial curriculum. The performances are evaluated by teachers using grade level achievement criteria. The data derived from these evaluations are used by the teachers as part of their assessment of the individual student's achievement and growth and are aggregated at the school and district level to provide district accountability information.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions-Reform/Innovation***
Elliott Asp

Littleton Public Schools, Colorado
The Role of Alternative Assessment in Systemic Change

This presentation will focus on the relationship of curriculum, instruction and assessment in providing energy and direction for restructuring at the school and district level in a medium-sized suburban school district. A school centered model for the assessment of student progress towards district level outcomes will be outlined. Examples of school and district assessment-driven reform efforts will be presented. These include performance-based high school graduation requirements, a middle-level exit outcomes and assessments project, performance assessment as curriculum and instruction, school and district reporting mechanisms, and district structures and procedures that promote, support, and sustain change at the school level.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Reform/Innovation***
Kevin Castner

Frederick County Public Schools, Maryland
School Reform and Innovative Assessment

The purpose of this presentation is to provide (a) a summary of the System for Effective Instruction—a process developed to support the Frederick County Public School System's outcome-based mission of "Teaching for Quality Learning for All Students"; and (b) an example of how authentic performance assessment drives this process. The system ensures an alignment of the written, taught and tested curriculum. It answers the question "What is it that we want our students to be effective at, and how will we assess that effectiveness?"

The essential components of the System for Effective Instruction are a three-level, essential curriculum model that includes course objectives, discipline goals and exit outcomes that are supported by a Criterion Referenced Evaluation System (CRES). The clear intent is to determine what is important for students to be able to do and how to measure it. The measurement, which will be done through formative and summative assessments, incorporates assessment into the everyday instructional process, thereby demystifying the standards and assessment process for staff, students and parents. CRES is performance-based, emphasizing thoughtful mastery of important tasks rather than thoughtless knowledge of isolated facts and skills.

Saturday, September 12, 1992
8:30 a.m. - 10:15 a.m. ***What We Know: Small Group Sessions–Reform/Innovation***
Jay McTighe

Maryland Assessment Consortium
The Maryland Assessment Consortium: A Collaborative Approach to Advancing Performance Assessment

Few states are moving as quickly and as comprehensively as the state of Maryland, with full implementation of summative performance measures in language arts, mathematics, social studies and science in elementary through middle school in the

spring of 1992, and in high school scheduled for the spring of 1994. Increasingly recognized is the extent to which assessment drives the instructional process. The link between quality assessment and quality instruction is the ability of teachers to adjust instruction based on the feedback provided by a series of formative assessments. While the Maryland State Department of Education has primary responsibility for the summative measures, local school districts must have access to parallel formative assessments so that this critical adjustment can occur.

The Maryland Assessment Consortium represents a unique collaboration of 24 school districts. Established in 1991, the Consortium exists to address two primary goals: (a) to design and deliver high-quality staff development programs on performance assessment for member districts; and (b) to collaboratively develop, field test, validate, and disseminate formative performance assessments for use by teachers as part of their instructional programs. These performance assessments are distributed in print form and will be incorporated into a computer "bank" for electronic dissemination. In this session, we'll describe the operation of the Consortium as a model for inter-district collaboration. We'll examine the staff development efforts as well as the process used in developing and validating performance assessment tasks. Finally, examples of performance assessments developed through the Consortium will be reviewed.